Elsevier Editorial System(tm) for Fisheries Research

Manuscript Draft

Manuscript Number: FISH9347R1

Title: Digital image analysis of flatfish bleeding injury

Article Type: Research Paper

Keywords: Algorithm; Automated image analysis; Haemorrhage; Observer bias

Corresponding Author: Dr. Sebastian Sebastian Uhlmann, PhD

Corresponding Author's Institution: Flanders Research Institute for Agriculture, Fisheries and Food (ILVO)

First Author: Sebastian Sebastian Uhlmann, PhD

Order of Authors: Sebastian Sebastian Uhlmann, PhD; Steven Verstockt, PhD; Bart Ampe, PhD

Manuscript Region of Origin: BELGIUM

Abstract: To improve accuracy of post-release mortality predictions and facilitate the routine collection of information about physical condition of catches after commercial fishing capture, traditional visual assessment by potentially subjective human observers or raters may be automated by digital image analysis. The purpose of this study was to develop a method and device that can eliminate subjectivity in scoring external injury of commercially beam-trawled flatfish by taking standardized, high resolution images to allow for automated calculation of the % surface area of visible bleeding injury relative to the whole fish based on digital image analysis. A reference library was compiled by photographing ventral sides of 67 fish of six flatfish species of different sizes and freshness (fresh vs defrosted). All fish were sourced from the R/V Simon Stevin while beam-trawling in the Belgian coastal zone of the Southern North Sea. All images that were neither over- nor under-exposed were compiled (n=51) and scored for the extent (%) of multifocal cutaneous petechial ('point bleeding'), and suffusion or haemorrhaging ('bruising') of the ventral head and body region, respectively by three experienced raters using a continuous scale (between 0 and 100%). Then, several state-of-the-art computer vision algorithms were tested on the dataset to develop a protocol that can 1) align each image; 2) identify fin, body and head regions; and 3) quantify the surface area of bleeding injury of each region by using appropriate thresholding techniques. For validation of the computer-derived % surface coverage estimates of bleeding injury, these were compared to the average rater's score. For bruising injury, a significant difference between human- vs computer-derived scores persisted. For point bleeding of the head region, computer-based estimates of % coverage were not different from those of the human raters. Overall, species, size and their freshness did not have a significant effect. By consistently recording the coverage of externally visible bleeding injury, this image analysis protocol may find its application in measuring the effect of different capture techniques on whole fish quality, and in improving vitality assessments as part of

the transition towards a more sustainable fishery and the implementation of the European Landing Obligation.

Research Data Related to this Submission

Title: Data for: Digital image analysis of flatfish bleeding injury
Repository: Mendeley Data
https://data.mendeley.com/datasets/swfg5fpzjn/draft?a=c1becedb-14e6-4bed-ab49-2997e08c0d7e

Abstract

To improve accuracy of post-release mortality predictions and facilitate the routine collection of information about physical condition of catches after commercial fishing capture, traditional visual assessment by potentially subjective human observers or raters may be automated by digital image analysis. The purpose of this study was to develop a method and device that can eliminate subjectivity in scoring external injury of commercially beam-trawled flatfish by taking standardized, high resolution images to allow for automated calculation of the % surface area of visible bleeding injury relative to the whole fish based on digital image analysis. A reference library was compiled by photographing ventral sides of 67 fish of six flatfish species of different sizes and freshness (fresh vs defrosted). All fish were sourced from the R/V *Simon Stevin* while beam-trawling in the Belgian coastal zone of the Southern North Sea. All images that were neither over- nor under-exposed were compiled (n=51) and scored for the extent (%) of multifocal cutaneous petechial ('point bleeding'), and suffusion or haemorrhaging ('bruising') of the ventral head and body region, respectively by three experienced raters using a continuous scale (between 0 and 100%). Then, several state-of-the-art computer vision algorithms were tested on the dataset to develop a protocol that can 1) align each image; 2) identify fin, body and head regions; and 3) quantify the surface area of bleeding injury of each region by using appropriate thresholding techniques. For validation of the computer-derived % surface coverage estimates of bleeding injury, these were compared to the average rater's score. For bruising injury, a significant difference between human- vs computer-derived scores persisted. For point bleeding of the head region, computer-based estimates of % coverage were not different from those of the human raters. Overall, species, size and their freshness did not have a significant effect. By consistently recording the coverage of externally visible bleeding injury, this image analysis protocol may find its application in measuring the effect of different capture techniques on whole fish quality, and

61 in improving vitality assessments as part of the transition towards a more sustainable fishery
62 and the implementation of the European Landing Obligation.
63
64 *Keywords:* Algorithm; Automated image analysis; Haemorrhage; Observer bias

65 **1. Introduction**

66 To evaluate the extent and severity of external bleeding injury of whole animals
67 captured by fishing, their condition can be visually assessed by observers (here 'raters') and
68 may be scored on a presence/absence (Davis and Ottmar 2006; Davis 2010; Uhlmann et al.
69 2016a), or categorical/ordinal scale (Main and Sangster, 1988; Esaiassen et al., 2013;
70 Meeremans et al., 2017). This can generate data to evaluate the effect of fishing gears and
71 operations on their (post-)capture condition (Suuronen et al., 1996; Kinds et al., 2015;
72 Soetaert et al., 2015). For example, the occurrence of multifocal petechial haemorrhages (i.e.,
73 point bleeding) and suffusion (i.e., bruising) of the head or body region were strong predictors
74 of delayed mortality of European plaice (*Pleuronectes platessa*) (Depestele et al., 2014, 2016;
75 Uhlmann et al., 2016a,b;). Haemorrhages may differ by the degree of how much blood had
76 infused surrounding tissue (Desender et al., 2016). To quantify the surface area covered by
77 bleeding injury, a rater typically scans the body for blood spots and bruises. Although cost
78 efficient and simple to perform, ratings typically involve an abstraction and interpretation of
79 criteria, which might be influenced by experience, cognitive ability, and other personal traits,
80 and hence might be biased (Meagher 2009, Burghardt et al., 2012; Tuyttens et al., 2014;
81 Meeremans et al., 2017).

82 Digital image analysis may offer an alternative to deliver more accurate and repeatable
83 results in a standardized procedure (Balaban et al., 2011; Davis and Ottmar, 2006; Kenney et
84 al., 2015). To quantify the amount of bleeding injury using normal white light (as opposed to
85 fluorescent or polarized light; Davis and Ottmar, 2006; Balaban et al. 2011, respectively),
86 Kenney et al., 2012 described procedures for North American yellowtail flounder (*Limanda*
87 *ferruginea*), focusing on detecting discolorations among photographed fillets as a quality
88 indicator. By using thresholding techniques to separate out injured tissue from the whole
89 fillet, a software-based algorithm was developed to draw outlines around the entire

circumference of the fillet and those areas that were bruised (Kenney et al., 2012). Difficulties were experienced to verify whether the detected area was a true bleeding rather than an artefact shading from the back light (Kenney et al., 2012).

In this study, the focus was on whole fish as opposed to fillets. The goal was to come up with a suite of digital image analysis techniques, which are widely applicable and which should be able to be used on-board of a (fishing or research) vessel in less optimal and under white-light conditions, to facilitate data collection during at-sea sampling surveys. Furthermore, depending on the time of impact (i.e., freshness of the injury), the thickness of the fish and the tissue depth it has occurred in, colours of bleeding may change and thus, may pose a challenge to accurate recognition. Thus, the algorithm should be able to cope with these issues and estimate/report its own certainty. Considering the above, in this study, digital colour photo images of various species of beam-trawled flatfish were digitally analysed to quantify the amount of bleeding injury on the ventral sides of the head, body and fin regions.

2.0 Material and methods

2.1 Data collection

In total, three trips were done with the R/V *Simon Stevin* equipped with a 3-m beam trawl and 32 mm codend mesh to collect fish to build a reference library of digital colour photographs. Sixty-six specimens from five species of flatfish (dab, *Limanda limanda*, *n*=24; European flounder, *Platichthys flesus*, *n*=4; European plaice, *Pleuronectes platessa*, *n*=24; common sole, *Solea solea*, *n*=8; Mediterranean scaldfish, *Arnoglossus laterna*, *n*=6; and turbot, *Scophthalmus maximus*, *n*=1) were collected and kept on ice or were frozen (Table 1). The treatment “fresh” vs “frozen” was introduced to evaluate whether it has an impact on the colour of the bleeding and requires different analytical techniques. Fish were photographed

using an existing laboratory set-up in Merelbeke, Belgium, comprising of a Nikon D200 dSLR camera (Nikon Corp., Tokyo, Japan) with a Nikon AF-S 18−200 mm F/3.5−5.6G ED-IF AF DX VR lens, a lighting table, PVC dome and top lights (Fig. 1). A grey card (Kontrollkarte, Novoflex Präzisiontechnik GmbH, Germany) was used to calibrate for the white balance of the images by taking a full-frame picture and metering using the auto modus of the white balance function of the camera. Before starting to take images of the fish, the camera was set to a manual modus to be able to choose shutter speed and aperture. From the ventral side of each fish a series of images was taken using different settings for shutter speed, aperture, light (with/without dome) (Online supplement). The ISO setting was kept at 100. In total, 267 pictures were taken and, of those, a selection of 51 pictures (those neither under- or over-exposed) were further used in this study. Exposure was determined based on a pixel light saturation intensity value between 180 and 200 (Fig. 2).

2.2 Digital image analysis

To quantify the surface area coverage of bleeding injury on the ventral side of flatfish, several techniques were evaluated to align and orient each image, segment the body into head, body and fin regions, and then finally detect bleeding injury of each of these regions. The first step of our automatic contour detection algorithm (written in Python) was segmenting the white-noisy background from the fish body. For the contour detection a combination of existing OpenCV techniques was used that included: i) contrast enhancement using CLAHE – Contrast Limited Adaptive Histogram Equalization (Zuiderveld, 2013); ii) combined edge/intensity based thresholding; iii) morphological filtering (opening and closing); and iv) connected component analysis (for blob filtering/selection). Then, by raytracing the fish's contour from its center pixel, calculated from the first order central moments of the fish's blob, the contour of the fish was detected. A blob is a region/group of pixels that share the

same properties/features. Raytracing implies that lines are projected at equally-spaced angles leading from the centre pixel to the outer edge of the fish's contour. For all angles, starting from 0 degrees, the length of each of these rays (i.e., the distance from the centre pixel to this edge point) was calculated and plotted. The contour points that corresponded to the peaks of the plots (i.e., the local maxima) form the head-tail line that is used to rotate the fish/image in the horizontal orientation. This orientation step facilitated the segmentation of the head, body and fin regions, which was the second step. K-means clustering of the pixel colors was evaluated to segment the fish into fin and body regions. A Gabor filter (Premana et al., 2018), and a novel masking technique based on gradient-based probability maps for each fish segment (trained on hand-labeled data) were used to improve the accuracy of the fin and head detection. The last step focused on detecting bleeding injuries and estimating the percentage of pixels in each region that represented these injuries equating to a % surface area of bleeding injury for each of the three regions: head, body and fins.

2.3 Validation

To validate the proposed automatic workflow of the digital image analysis techniques, three experienced raters independently scored each of the 51 images for the % ventral surface covered by point bleeding and bruising of the head and body regions, respectively; using a continuous scale between 0 and 100% (following Meeremans et al., 2017). The raters did not score damage to the fins, because it was considered to be too difficult to do this properly from photos without handling the fish when fins can be fanned out for improved visibility and assessment. The same limitation applies to score fin damage based on digital image analysis. The relative and absolute differences in a given human rater's and computer-derived score (from automated digital image analysis; 'auto') from the rater's average was graphically explored. To test for significant differences among averaged rater's and auto scores a linear-

mixed model (LMM) was used with an interaction between source (i.e., rater vs auto) and type (i.e., four injury types: bruising or point bleeding of the head or body region), treatment (fresh vs defrosted) and species (i.e., dab, plaice, sole, flounder, and scaldfish) as fixed effects and image ID as random effect. Images of turbot were excluded due to a lack of replication for this species. To test whether the differences between the auto and averaged rater scores became smaller, when scores from a given rater were excluded from the data, a generalized linear model was used to test whether the relative difference between auto and the average score from two rater's was significantly different as a function of injury types. Finally, to validate the scores from the digital image analysis and to test the alternative hypothesis that the auto scores are similar to the mean of the human rater scores an equivalence test was done between the scores of the human raters (defined as treatment) and the auto scores (control), using the above LMM model. Equivalence was defined, when the auto score deviated less than 10% from the average of the three rater's score.

3.0 Results

3.1 Digital image analysis

3.1.1 Contour detection

The combination of the above mentioned techniques resulted in the highest contour detection accuracy compared with either technique used in isolation. It resulted in a binary black and white image, in which the white region was the region of interest (ROI; i.e., the fish), and the black region was the background (Fig. 3). Based on all pixel locations of the white blob, the ROI was generated by taking the minimum and maximum coordinates in both directions. The resulting ROIs were also shown in Fig. 3 (blue rectangular bounding boxes). Plotting the distance of each ray (traced from the centre pixel to the contour line; raytracing

technique, as explained above) resulted in distinctly shaped curves (Fig. 4). The contour peaks indicated the tail and the head regions (Fig. 4). The highest peak was (in most cases) the tail. If for some species of fish this would represent the head (or another body part), the algorithm can be easily adapted to take this into account. The results of the orientation correction are shown in Fig. 5. In these examples, the fish were always oriented with its tail to the left and its head to the right.

3.1.2 Segmentation

Color-based K-means clustering generated robust fin segmentation results for fishes with fins that were much brighter than the body of the fish. When the input of the K-means clustering was changed to the Gabor-filter responses of a Gabor bank that focused on different types of texture patterns, fin segmentation improved across all fish species (Fig. 6). Unfortunately, a similar approach for head detection did not yield satisfying results. Instead as a proxy, results from the raytracing were used. Based on the size of the fish (from its blob statistics), a head ROI was scaled around the nose (Fig. 7). In the current version of the software we use a fixed scale factor of 15%. This scale factor, however, will be species-specific, and will be defined in future work by expert labeling of a representative dataset for each type of fish.

Another approach to more accurately detect the head region of the fish was to extend the Gabor-filter based K-means clustering and to take into account the probability of each pixel belonging to the fin or head regions based on its pixel location in the correctly oriented image. For both the head and the fin an RGB mask was constructed (Fig. 8). Species-specific masks were constructed based on the training set. The mask was mapped onto the ROI of the fish

211 and corresponding mask values were used as an additional feature for each pixel. The results
212 of this approach were very satisfactory (Fig. 9).

213

214 *3.1.3 Bleeding injury detection*

215 A simple (yet not so efficient) solution to detect any bleeding injury consisted of
216 converting the Red-Green-Blue (RGB) color image to the Hue-Saturation-Value (HSV) color
217 space and then to filter out the 'injury pixels' using two thresholds in the upper and lower red
218 zones of the Hue channel. Hue is a color appearance parameter. Next, based on the size of
219 connected injury pixels the injury was classified as either a point bleeding or bruising (Fig.
220 10). In combination with the head, body and fin segmentation, the percentage of injury pixels
221 in each region was estimated.

222

223 *3.2 Validation*

224 Overall, the computer-derived auto scores were similar to the human rater scores for point
225 bleedings, but deviated for bruisings. More specifically, the relative and absolute differences
226 in a given human rater's compared to the average score of all three raters showed that while
227 raters 1 and 2 either under- or overestimated bruising injuries, respectively , rater 3 scored
228 closest to the average of all three rater's scores (Fig. 11a). The computer-derived scores (from
229 automated digital image analysis) ranged closest to those of rater 1 (-18.0% vs -7.8% and -
230 10.0% vs 6.6% for bruising body and head for computer-vs human-derived scores,
231 respectively; Fig. 11a); and rater 3 scored closest to the average of all three raters (0% and
232 5.8% for bruising body and head; Fig. 11a). There was a significant interaction between rater
233 and injury type (LMM; d.f. = 9, $p < 0.001$). When testing for differences among averaged
234 rater scores and auto scores using the linear-mixed model (LMM), only the interaction

between source (rater vs auto) and injury type was included in the final model, because both fixed effects for species and treatment (fresh vs defrosted) were not significant. The results showed that apart from point bleeding to the head, there were significant differences between the averaged rater and auto score for all other injury types (Table 2). Leaving out the scores from either rater 1 or 3, resulted in some improvement to minimize the differences between auto and the average of the other two raters scores for all injury types apart from body bruising (Table 3).

Finally, from the equivalence test results it was concluded that only for point bleeding of the head region, the auto scores fell within an arbitrary 10% deviation from the mean of the averaged raters scores (Table 4; Fig. 12). For the other injury types, the alternative hypothesis of equivalence could not be accepted (Table 4; Fig. 12). It was noted, however, that out of 600 scores produced by all the three raters, zero bleeding was observed 7 and 23 times for bruising of the body and head, vs 74 and 124 times for point bleeding of the body and head, respectively. Out of 200 computer-derived scores, zero bleeding was registered 17 and 18 times for bruising of the body and head, vs 20 and 5 times for point bleeding of the body and head.

4.0 Discussion

Digital image analysis techniques can be a useful tool to objectively quantify external injury of beam-trawled-and-discarded flatfish. Here, we developed low computational-cost techniques (< 1 second on average on a standard CPU) that rendered promising results and were simple to apply under normal, white-light conditions. These methods did not require any reagents to treat animals prior to being photographed as opposed to presumptive techniques (e.g., Colotelo et al., 2009) other than simply having a fishes surface cleaned from any debris

before taking pictures off them. Apart from accurately drawing an outline around areas that were marked by bleeding, the proposed methodology included techniques that can facilitate species recognition (e.g., based on the raytracing graphs features), which is an important area of interest to improve the full and automated documentation of fisheries catches (White et al., 2006; Mangi et al., 2013; French et al., 2015, 2019; Mortensen et al., 2016).

Despite such potential, the validation and equivalence with observational rater scores has to be demonstrated before the technique can be used to complement data collection programmes by observers. There was variation among the three rater's ability to agree upon extent of bleeding injury despite their practical prior experience; and there were differences between averaged human rater's and computer-derived scores (from automated digital image analysis), especially for bruising injuries of both the head and body regions. For a future validation it will be important to increase the sample size of assessed images, include a training component for raters to agree upon scoring criteria prior to commencing with a scoring exercise and also allow a validation across a full range in the data. In this case, gear deployment duration may have been too short to induce point bleeding injury.

However, the proposed methodology still has some limitations in its ability to segment the fish into regions and in its efficiency in delineating areas of bleeding injury. This is relevant, because of the potential, differential impact of bleeding injury and its locality on welfare and survivability (Desender et al., 2016; Uhlmann et al., 2016a, 2016b). To improve segmentation, depending on the species of fish, other types of Gabor filters may be needed for the segmentation by head, body and fin regions. The focus thereby should be on selecting least complex techniques (i.e., low computational costs). This can possibly be learned during a training phase using a larger reference picture dataset and by comparing the results from both approaches. Secondly, to improve the agreement among surface area estimates of bleeding injury between raters versus automated, digital image analysis, the head region

detection may need to be optimized (currently, it is estimated to be larger than what raters considered as head region). The definition of the head regions needs to be made by the raters. However, even though that based on the computer segmentation a larger head region was generally defined, overall the mean auto scores were smaller than those from the raters (Fig. 12). Further, species-based histogram matching may be used in a future study as a pre-processing step so that the input images' histogram matches a specified histogram that determines more appropriate thresholds. To improve distinctions between bruising and point bleeding, which is currently based on blob size, an optimization routine may be required to match what raters define as point bleeding/bruising – a classifier can then be trained upon these data to take decisions based on the insights gathered from the ground truth labeled blob features. Future work should also evaluate whether the raytracing contour description can be reliably used for species identification, which in turn could determine the parameter selection for several, subsequent building blocks of the architecture that is proposed in this study, e.g. when we know the type of fish, the head/fin/body segmentation can be trained for each type specific. It could also be interesting to study feature learning based approaches, e.g. using convolutional neural networks (see Mangi et al., 2013; French et al., 2019) and evaluate if they can achieve higher accuracy as the proposed engineering-based workflow.

Digital image analysis, combined with machine learning techniques, has great potential in fisheries science to facilitate data collection about animal welfare and fishing impact indicators and is not restricted to food processing and market quality evaluations, amongst other applications (White et al., 2006; ICES, 2019; van Helmond et al., 2019). When eventually operationalizing *in situ*, on-board applications, valuable information may be gathered about the condition of catches that could benefit seafood certification schemes such as the Marine Stewardship Council (MSC) and in general contribute to more sustainable fishing operations.

Acknowledgements

We would like to thank the crew of the R/V *Simon Stevin* and staff from the Flanders Marine Institute for rigging up the beam trawl and helping us to collect the fish. This work was supported by the European Maritime and Fisheries Fund (Grant Number: LV-01-090415). We are grateful to Ryan de Cloedt and Thijs Decoene for data collection and building the reference library of digital photographs, following expert advice by Flanders Research Institute of Agriculture, Fisheries and Food colleagues Peter Lootens and Simon Kool. This work would not have been possible without the input from all students who participated in the Multimedia class of Steven Verstockt at the University of Ghent (campus Kortrijk) for their groundwork in establishing the appropriate procedures. The comments received by the journal editor and two reviewers were very much appreciated.

References

Balaban, M.O., Ünal Şengör, G.F., Soriano, M.G., Ruiz, E.G., 2011. Quantification of gaping, bruising, and blood spots in salmon fillets using image analysis. J. Food Sci. 76. https://doi.org/10.1111/j.1750-3841.2011.02060.x

Burghardt, G.M., Bartmess-Levasseur, J.N., Browning, S.A., Morrison, K.E., Stec, C.L., Zachau, C.E., Freeberg, T.M., 2012. Perspectives - Minimizing Observer Bias in Behavioral Studies: A Review and Recommendations. Ethology 118, 511–517. https://doi.org/10.1111/j.1439-0310.2012.02040.x

Colotelo, A.H., Cooke, S.J., Smokorowski, K.E., 2009. Application of forensic techniques to enhance fish conservation and management: Injury detection using presumptive tests for blood. Endanger. Species Res. 9, 169–178. https://doi.org/10.3354/esr00178

Davis, M.W., Ottmar, M.L., 2006. Wounding and reflex impairment may be predictors for mortality in discarded or escaped fish. Fish. Res. https://doi.org/10.1016/j.fishres.2006.09.004

Davis, M.W., 2010. Fish stress and mortality can be predicted using reflex impairment. Fish Fish. 11, 1–11. https://doi.org/10.1111/j.1467-2979.2009.00331.x

Depestele, J., Desender, M., Benoît, H.P., Polet, H., Vincx, M., 2014. Short-term survival of discarded target fish and non-target invertebrate species in the "eurocutter" beam trawl fishery of the southern North Sea. Fish. Res. 154, 82–92. https://doi.org/http://dx.doi.org/10.1016/j.fishres.2014.01.018

Desender, M., Chiers, K., Polet, H., Verschueren, B., Saunders, J.H., Ampe, B., Mortensen, A., Puvanendran, V., Decostere, A., 2016. Short-term effect of pulsed direct current on various species of adult fish and its implication in pulse trawling for brown shrimp in the North Sea. Fish. Res. 179, 90–97. https://doi.org/10.1016/j.fishres.2016.02.018

Esaiassen, M., Akse, L., Joensen, S., 2013. Development of a Catch-damage-index to assess the quality of cod at landing. Food Control 29, 231–235. https://doi.org/10.1016/j.foodcont.2012.05.065

French, G., Mackiewicz, M., Fisher, M., Holah, H., Kilburn, R., Campbell, N., Needle, C., 2019. Deep neural networks for analysis of fisheries surveillance video and automated monitoring of fish discards. ICES J. Mar. Sci. https://doi.org/10.1093/icesjms/fsz149

French, G., Fisher, M., Mackiewicz, M., Needle, C., 2015. Convolutional Neural Networks for Counting Fish in Fisheries Surveillance Video. Proc. Mach. Vis. Anim. their Behav. 7.1-7.10. https://doi.org/10.5244/C.29.MVAB.7

ICES, 2019. Working Group on Machine Learning in Marine Science (WGMLEARN). ICES Scientific Reports. 1:45. 13 pp. http://doi.org/10.17895/ices.pub.5539

Kenney, J.L., Rahman, T., Manuel, H., Winger, P.D., 2015. Bruising patterns in commercially harvested yellowtail flounder (*Limanda ferruginea*). Fish. Res. 172, 79–84. https://doi.org/10.1016/J.FISHRES.2015.06.026

Kinds, A., Sys, K., Schotte, L., Mondelaers, K., Polet, H., 2015. VALDUVIS: An innovative approach to assess the sustainability of fishing activities. Fish. Res. https://doi.org/10.1016/j.fishres.2015.10.027

Main, J., Sangster, G.I., 1988. Scale damage and survival of young gadoid fish escaping from the cod-end of a demersal trawl, in: DeAlteris, J.T. (Ed.), Proceedings of Gear Selectivity and Survivability Workshop. University of Rhode Island Sea Grant Advisory Service, Narragansett, pp. 17–33.

Mangi, S.C., Dolder, P.J., Catchpole, T.L., Rodmell, D., de Rozarieux, N., 2013. Approaches to fully documented fisheries: practical issues and stakeholder perceptions. Fish Fish. n/a-n/a. https://doi.org/10.1111/faf.12065

Meeremans, P., Yochum, N., Kochzius, M., Ampe, B., Tuyttens, F.A.M., Uhlmann, S.S., 2017. Inter-rater reliability of categorical versus continuous scoring of fish vitality: Does it affect the utility of the reflex action mortality predictor (RAMP) approach? PLoS One 12. https://doi.org/10.1371/journal.pone.0179092

Meagher, R.K., 2009. Observer ratings: Validity and value as a tool for animal welfare research. Appl. Anim. Behav. Sci. 119, 1–14. https://doi.org/10.1016/j.applanim.2009.02.026

Mortensen, L.O., Ulrich, C., Olesen, H.J., Bergsson, H., Berg, C.W., Tzamouranis, N., Dalskov, J., 2017. Effectiveness of fully documented fisheries to estimate discards in a participatory research scheme. Fish. Res. https://doi.org/10.1016/j.fishres.2016.11.010

Premana, A., Wijaya, A.P., Soeleman, M.A., 2018. Image segmentation using Gabor filter and K-means clustering method, in: Proceedings - 2017 International Seminar on Application for Technology of Information and Communication: Empowering Technology for a Better Human Life, ISemantic 2017. pp. 95–99. https://doi.org/10.1109/ISEMANTIC.2017.8251850

Soetaert, M., Decostere, A., Polet, H., Verschueren, B., Chiers, K., 2015. Electrotrawling: a promising alternative fishing technique warranting further exploration. Fish Fish. 16, 104–124. https://doi.org/10.1111/faf.12047

Suuronen, P., PerezComas, J.A., Lehtonen, E., Tschernij, V., 1996. Size-related mortality of herring (*Clupea harengus* L) escaping through a rigid sorting grid and trawl codend meshes. ICES J. Mar. Sci. 53, 691–700.

Tuyttens, F.A.M., de Graaf, S., Heerkens, J.L.T., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van Laer, E., Ampe, B., 2014. Observer bias in animal behaviour research: Can we believe what we score, if we score what we believe? Anim. Behav. 90, 273–280. https://doi.org/10.1016/j.anbehav.2014.02.007

Uhlmann, S.S., Theunynck, R., Ampe, B., Desender, M., Soetaert, M., Depestele, J., 2016a. Injury, reflex impairment, and survival of beam-trawled flatfish. ICES J. Mar. Sci. 73, 1244–1254. https://doi.org/10.1093/icesjms/fsv252

Uhlmann, S. S., Theunynck, R., Ampe, B., Verkempynck, R., Miller, D. C. M., van Marlen, B., van der Reijden, K., Molenaar, P., Vanderperren, E., Polet, H. 2016b. Overleving door boomkor gevangen pladijs – survival of beam-trawled European plaice (*Pleuronectes platessa*). ILVO Mededeling 210. Flanders Research Institute for Agricultural, Fisheries and Food (ILVO), Oostende, Belgium. 172 pp.

van Helmond, A.T.M., Mortensen, L.O., Plet-Hansen, K.S., Ulrich, C., Needle, C.L., Oesterwind, D., Kindt-Larsen, L., Catchpole, T., Mangi, S., Zimmermann, C., Olesen, H.J., Bailey, N., Bergsson, H., Dalskov, J., Elson, J., Hosken, M., Peterson, L., McElderry, H., Ruiz, J., Pierre, J.P., Dykstra, C., Poos, J.J., 2019. Electronic monitoring in fisheries: Lessons from global experiences and future opportunities. Fish Fish. https://doi.org/10.1111/faf.12425

White, D.J., Svellingen, C., Strachan, N.J.C., 2006. Automated measurement of species and length of fish by computer vision. Fish. Res., 203–210 https://doi.org/10.1016/j.fishres.2006.04.009

Zuiderveld, K., 2013. Contrast Limited Adaptive Histogram Equalization, in: Graphics Gems. https://doi.org/10.1016/b978-0-12-336156-1.50061-6

List of tables

Table 1. Overview of the number of individuals and their treatment (kept fresh on ice after beam-trawl capture vs defrosted) per species of flatfish that were collected and eventually scored. /, if exposures of pictures were not of sufficient quality, or sample size too small, they were not considered for further analysis.

Species	**Scientific name**	**Treatment**	**Sample size (total)**	**Scored by humans & computer**
Common dab	*Limanda limanda*	Defrosted	6	6
		Fresh	18	17
European plaice	*Pleuronectes platessa*	Defrosted	4	4
		Fresh	20	15
Common sole	*Solea solea*	Defrosted	3	/
		Fresh	5	3
European flounder	*Platichthys flesus*	Fresh	4	3
Mediterranean scaldfish	*Arnoglossus laterna*	Fresh	6	2
Turbot	*Scophthalmus maximus*	Fresh	1	/

Table 2. Results from the linear-mixed model (LMM) testing for significant differences among averaged rater vs automated scores as a function of an interaction between source (rater vs auto) and type of injury (species and treatment were not significant, and thus excluded as fixed effects from the final model). Post-hoc Tukey test results of pairwise comparisons between rater and auto scores and their estimated marginal means (Emmean ± SE) and lower and upper 95% confidence intervals (CI) per bleeding injury types (bruising or point bleeding of the head or body regions, respectively). Groups with the same letter are not significantly different at p=0.05.

Injury	Source	Emmean	SE	df	lower.CL	upper.CL	Group
Bruising (body)	auto	7,532	1,67	269	4,24	10,83	a
	rater	28,173	1,66	266	24,9	31,45	b
Bruising (head)	auto	13,261	1,67	269	9,97	16,56	a
	rater	25,78	1,66	266	22,51	29,05	b
Point bleeding (body)	auto	1,01	1,67	269	-2,28	4,3	a
	rater	4,707	1,66	266	1,43	7,98	b
Point bleeding (head)	auto	3,074	1,67	269	-0,22	6,37	a
	rater	0,793	1,66	266	-2,48	4,07	a

433 **Table 3.** Results from the generalized linear model to test whether the relative difference
434 between auto and the average score from two rater's was significantly different as a function
435 of injury types, with the estimated marginal means (Emmean ± SE) of the difference and
436 lower and upper 95% confidence intervals (CI) per bleeding injury types (bruising or point
437 bleeding of the head or body regions, respectively). Groups with the same letter are not
438 significantly different at p=0.05.

439

Difference	Type	Emmean	SE	df	lower.CL	upper.CL	Group
Without rater 1	Bruising (body)	-19,67	2,01	53	-23,70	-15,63	1
	Bruising (head)	0,00	3,49	53	-6,99	6,99	2
	Point bleeding (body)	0,71	1,32	53	-1,93	3,36	2
	Point bleeding (head)	2,89	0,83	53	1,21	4,56	2
Without rater 2	Bruising (body)	-12,00	1,25	55	-14,50	-9,50	1
	Bruising (head)	-5,50	1,53	55	-8,56	-2,44	2
	Point bleeding (body)	0,71	0,82	55	-0,92	2,35	3
	Point bleeding (head)	2,89	0,52	55	1,85	3,92	3
Without rater 3	Bruising (body)	-19,14	1,52	56	-22,18	-16,10	1
	Bruising (head)	2,00	2,32	56	-2,65	6,65	2
	Point bleeding (body)	0,40	1,04	56	-1,68	2,48	2
	Point bleeding (head)	2,89	0,68	56	1,53	4,25	2

440

441

Table 4. Post-hoc contrasts and their confidence limits for the differences between the averaged rater's score (treatment) and auto (control) (used as an equivalence test).

Injury	Contrast	Estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Bruising (body)	auto - rater	-20,64	1,5	193	-24	-18	-14	0,00
Bruising (head)	auto - rater	-12,52	1,5	193	-15	-10	-8	0,00
Point bleeding (body)	auto - rater	-3,70	1,5	193	-7	-0,747	-2	0,01
Point bleeding (head)	auto - rater	2,28	1,5	193	-1	5	2	0,13

449 **List of figures**

450

451 **Fig. 1.** Laboratory-based photo studio set-up, comprising of a Nikon dSLR camera, a lighting
452 table, and top lights. A milk-glass coloured PVC dome is not pictured.

453

454 **Fig. 2.** A representative sub-set of images from the reference library (*n*=267 images) showing
455 multiple (between 3 and 4) images of a fish, taken with different exposure/lighting settings.
456 The image ID imprinted underneath each picture corresponds with the ID in the online
457 supplement data file and can be used to retrieve the exact camera settings. The annotated
458 dataset (*n*=51 images) contained pictures that were neither under- nor over-exposed. Each of
459 these images was scored independently by three experts for their extent of bleeding injury,
460 and also analysed by an automated digital image analysis. The following species are depicted
461 as per image ID: 2226, 2230, 2242, 2258, 2460, 2462, 2464, and 2467 = plaice; 2274, 2287,
462 2291, and 2306 = dab; 2402 = flounder; 2427 = sole; and 2455 = turbot.

463

464 **Fig. 3.** Fish segmentation (black-and-white image), contour detection (shown in red) and
465 region of interest (ROI) selection (shown in blue as rectangular box) on input images.

466

467 **Fig. 4.** Raytracing of a fish contour from the center of the white blob (center pixel). The
468 resulting contour graph with distinct peaks for the tail and head regions may be useful to
469 determine the species of fish and the orientation of the image.

470

Fig. 5. Automatic orientation correction (examples) based on proposed raytracing head-tail detection algorithm.

Fig. 6. Based on specific tissue texture patterns, a Gabor-filter based K-means clustering technique was used to segment fins from the rest of a fish body.

Fig. 7. Head region estimation (defined by the blue rectangle drawn around the head) using nose detection (from the ray tracing contour description) and a head region-of-interest scaling that is proportional to the size/area of the fish. The red colored zones are the detected bleeding/bruising of the fish. To illustrate the concept of head region estimation, three different individuals of plaice were pictured here.

Fig. 8. Probability masks used for the detection of a fishes' head (top) and fin (bottom) regions with their corresponding mask overlays. Red means the pixels very likely belonged to the fin/head, green it likely belonged to it and blue that it very unlikely belonged to it.

Fig. 9. Results of the proposed head, body and fin segmentation after overlaying the Gabor-filter and masking based k-means clusters, whereby the colours indicate the assigned body regions: blue=fins; green=body; and red= head). Results can possibly be further improved by some additional post-processing optimizations/filtering (e.g., head area is sometimes discontinuous).

Fig. 10. Bleeding injury detection. Injury pixels were shown in blue. Based on the size of the connected injury pixels the injury was classified as either a point bleeding or bruising.

Fig. 11. Relative (A) and absolute differences (B) of a given score (automated digital image analysis score – very dark grey, rater 1 – dark grey; rater 2 – white; rater 3 – light grey) from the average rater score between the 10^{th} and 90^{th} percentile. Line in the middle is the median.

Fig. 12. Results from the equivalence test plotting the estimated marginal mean (black dot) and its 95% confidence interval (grey rectangle) for the auto and averaged three rater scores per injury type (bruis_B – bruising [body]; bruis_H – bruising [head]; pnt_bld_B – point bleeding [body]; pnt_bld_H – point bleeding [head]).

Figure
Click here to download high resolution image

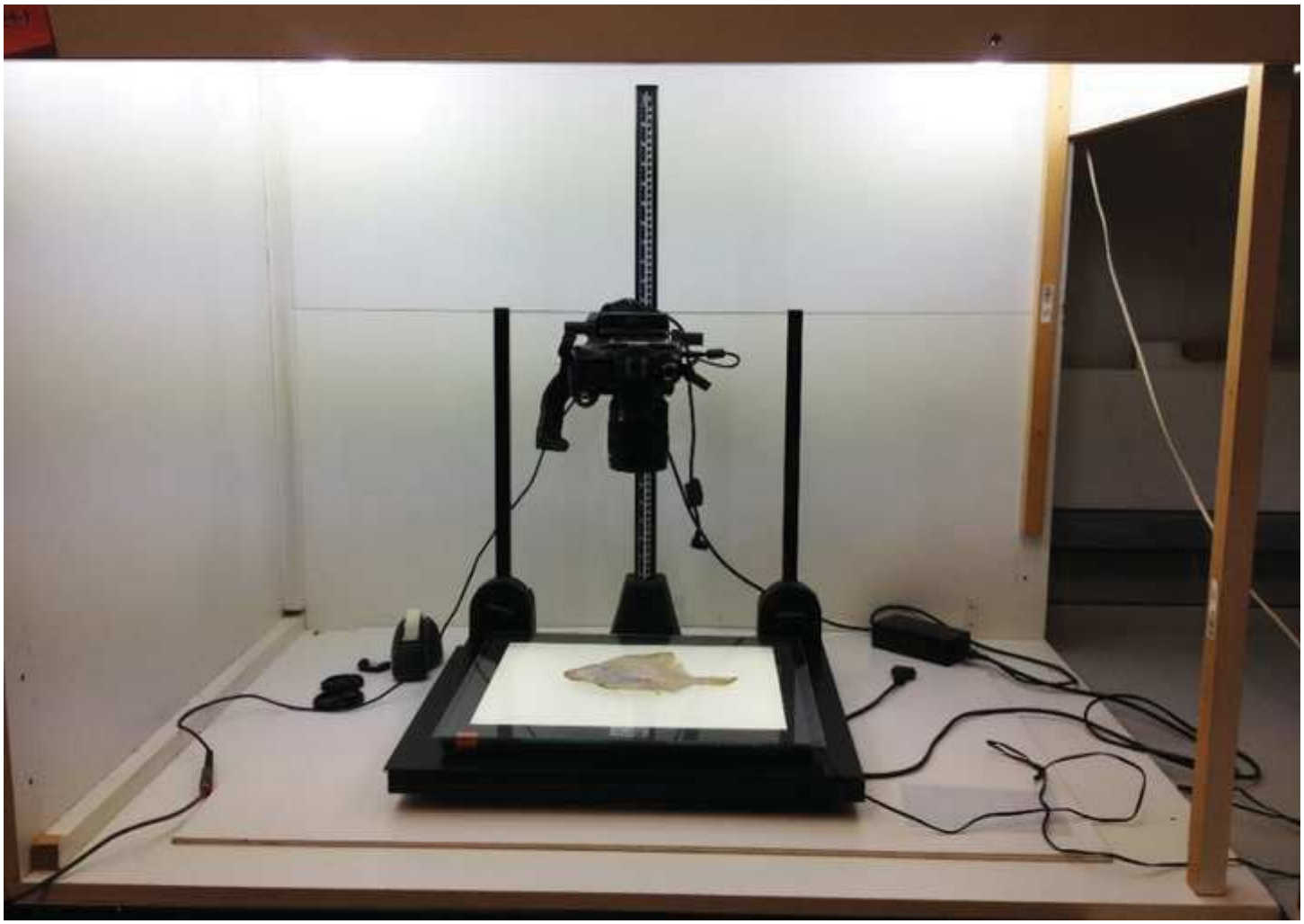

Figure 2
Click here to download high resolution image

Figure 3
Click here to download high resolution image

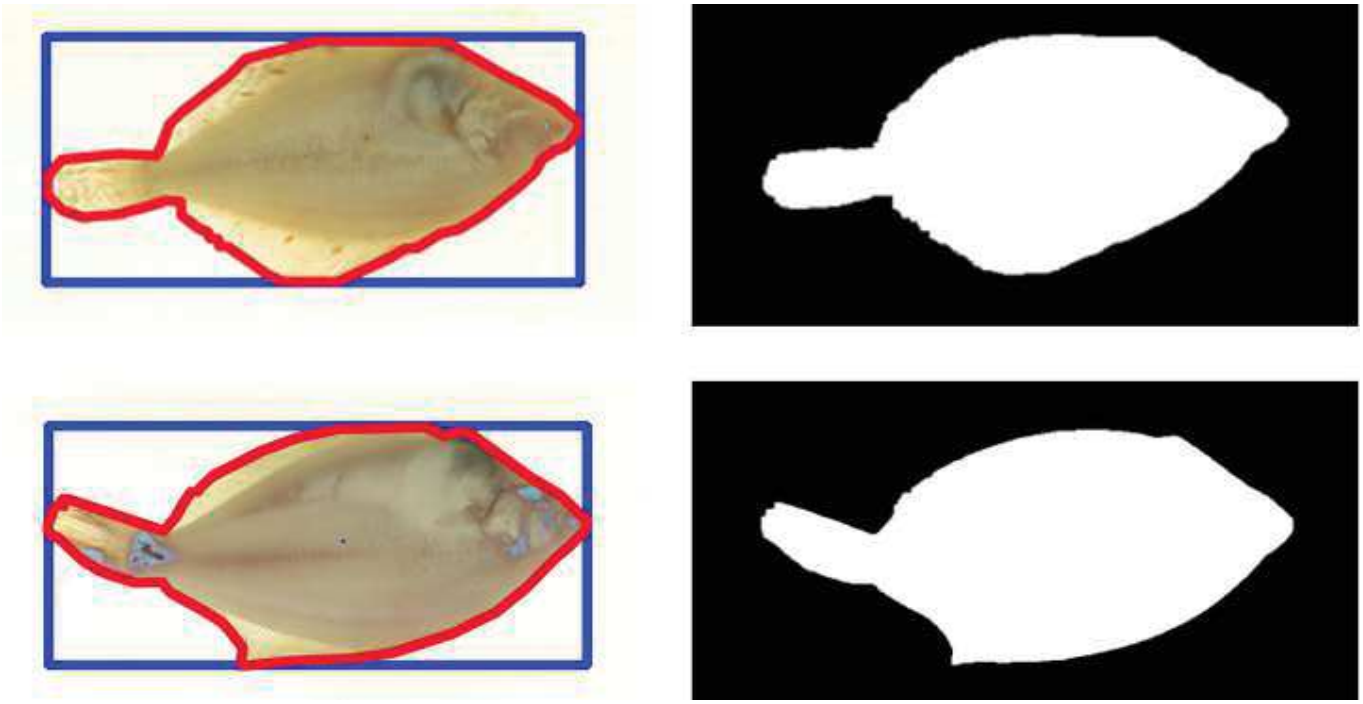

Figure
Click here to download high resolution image

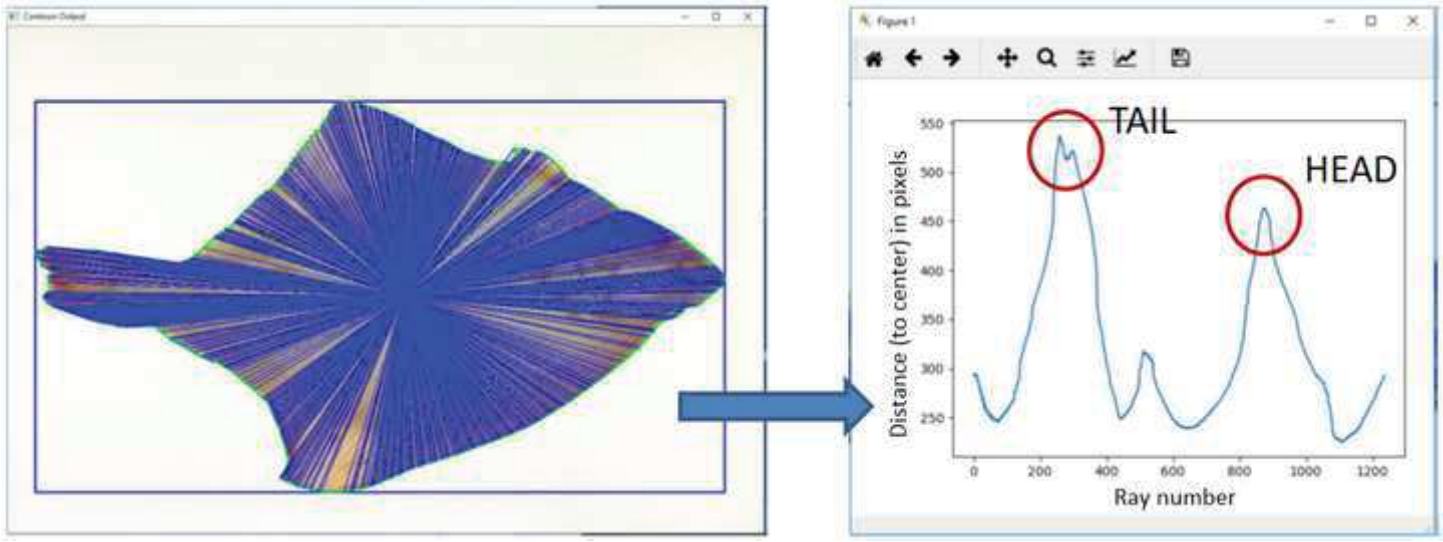

Figure
Click here to download high resolution image

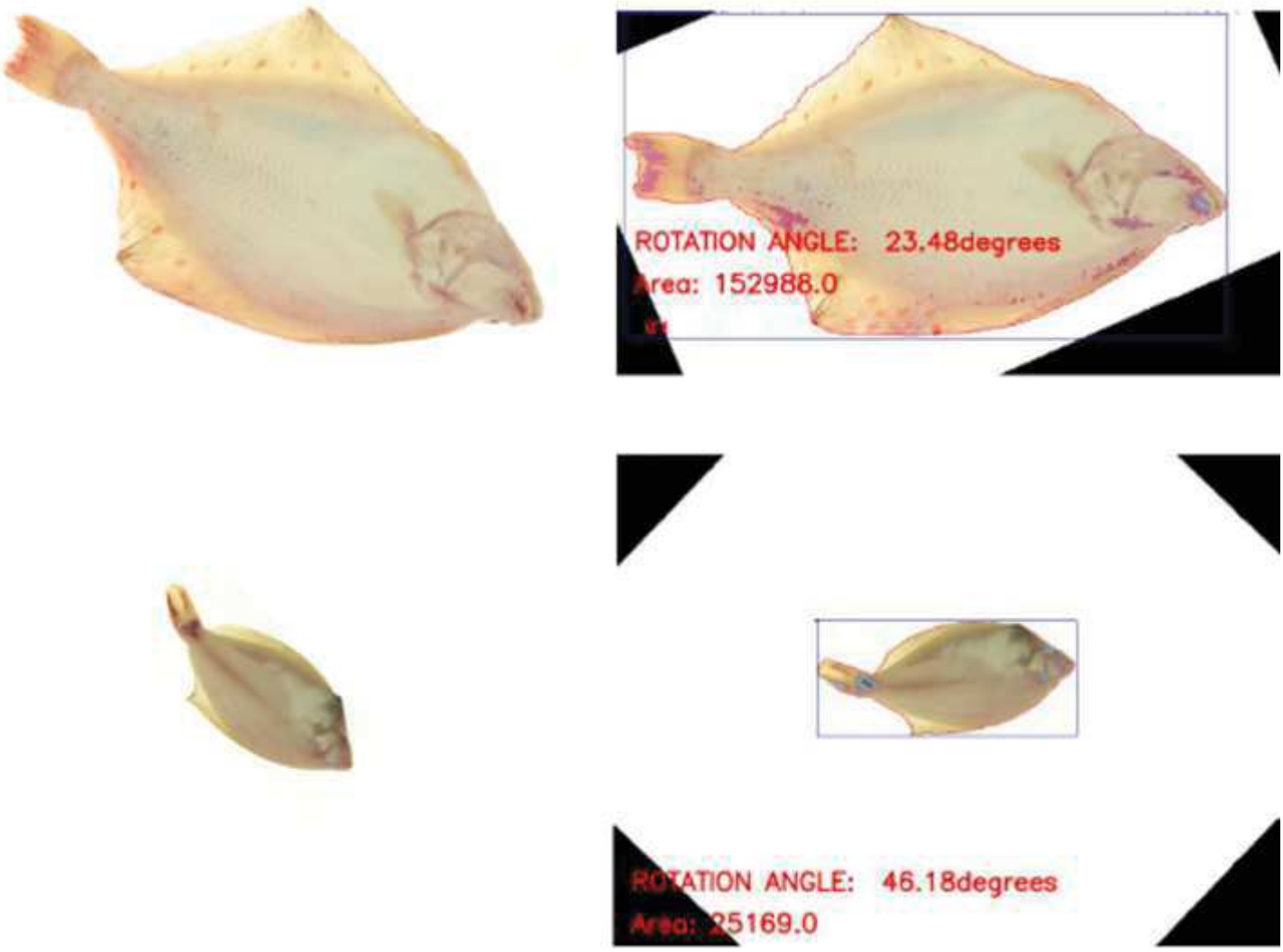

Figure 6
Click here to download high resolution image

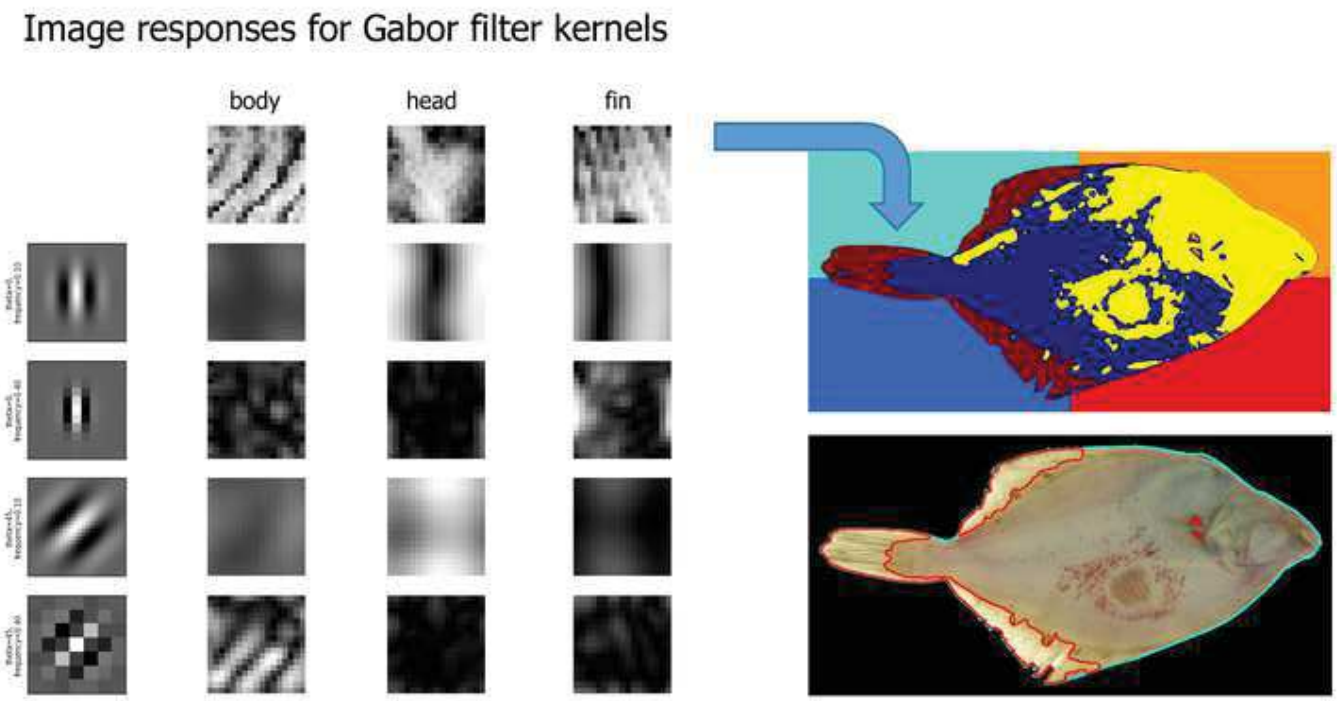

Figure
Click here to download high resolution image

Figure 8
Click here to download high resolution image

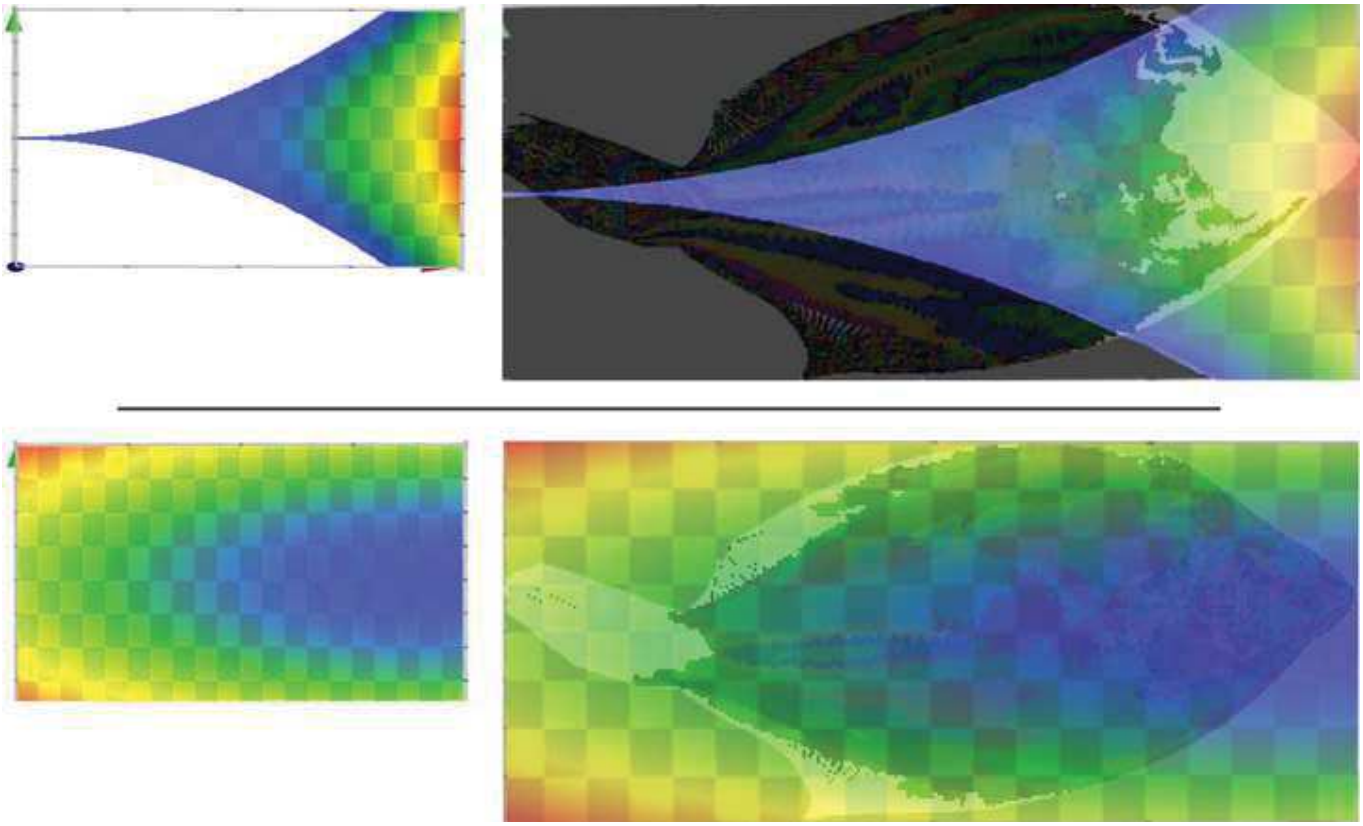

Figure
Click here to download high resolution image

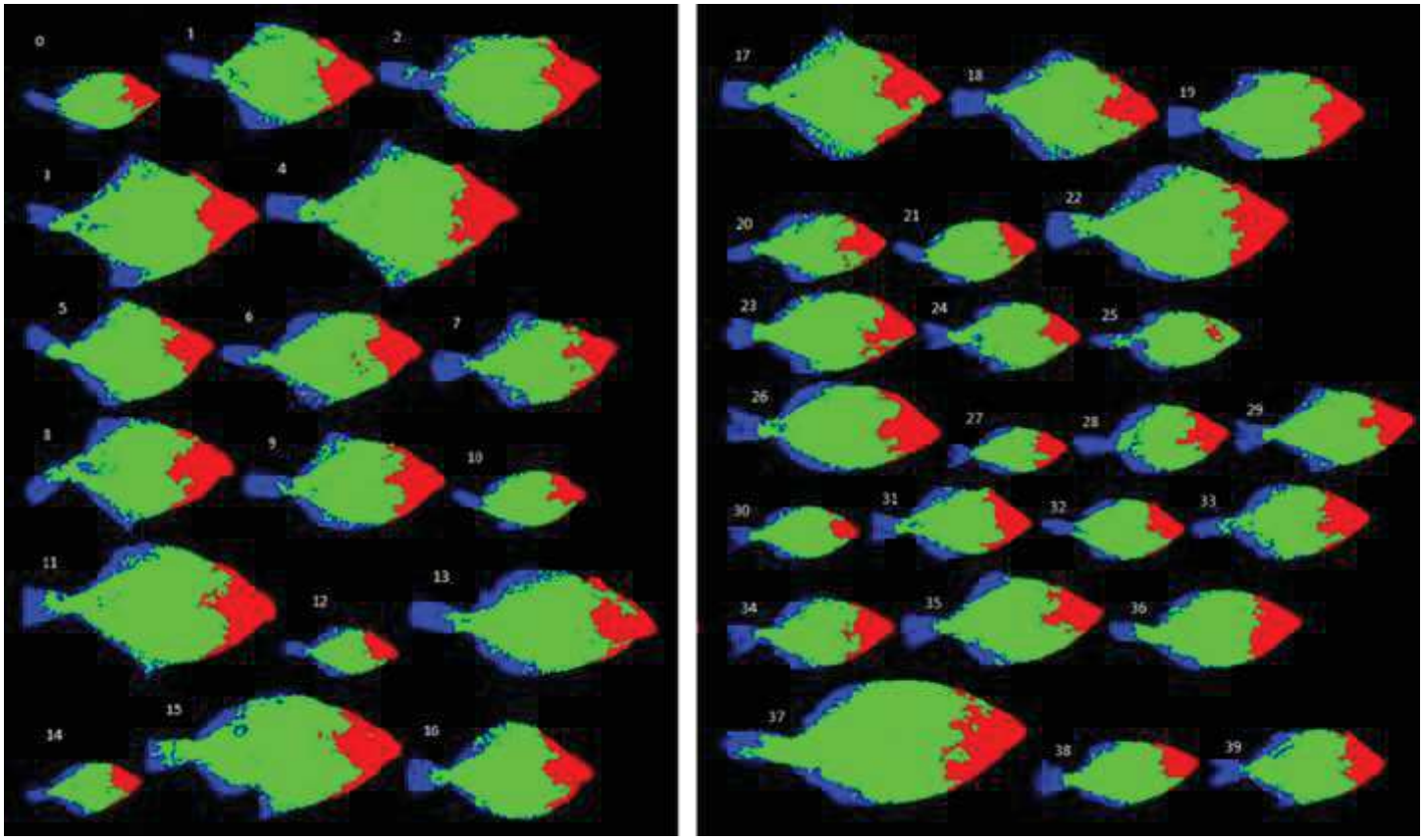

Figure
Click here to download high resolution image

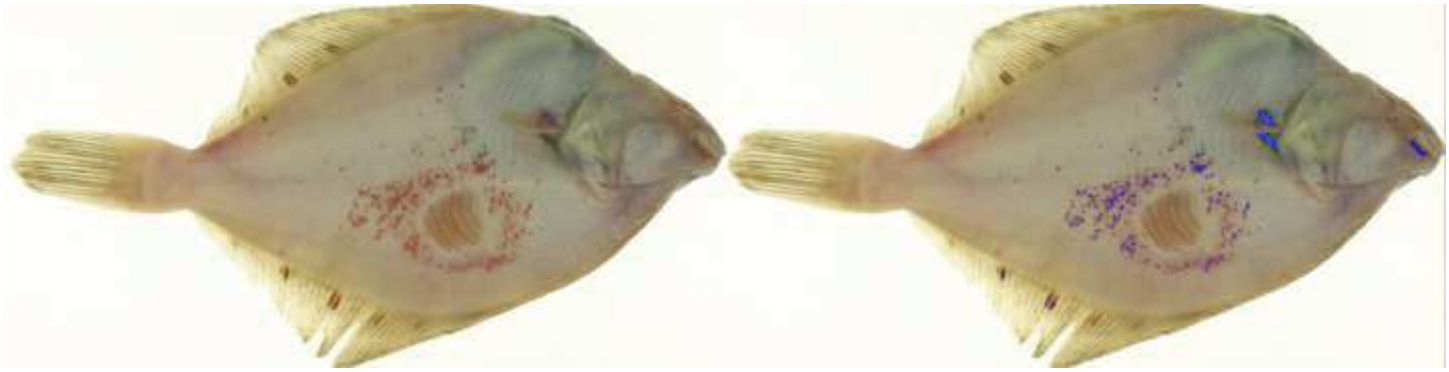

Figure
Click here to download high resolution image

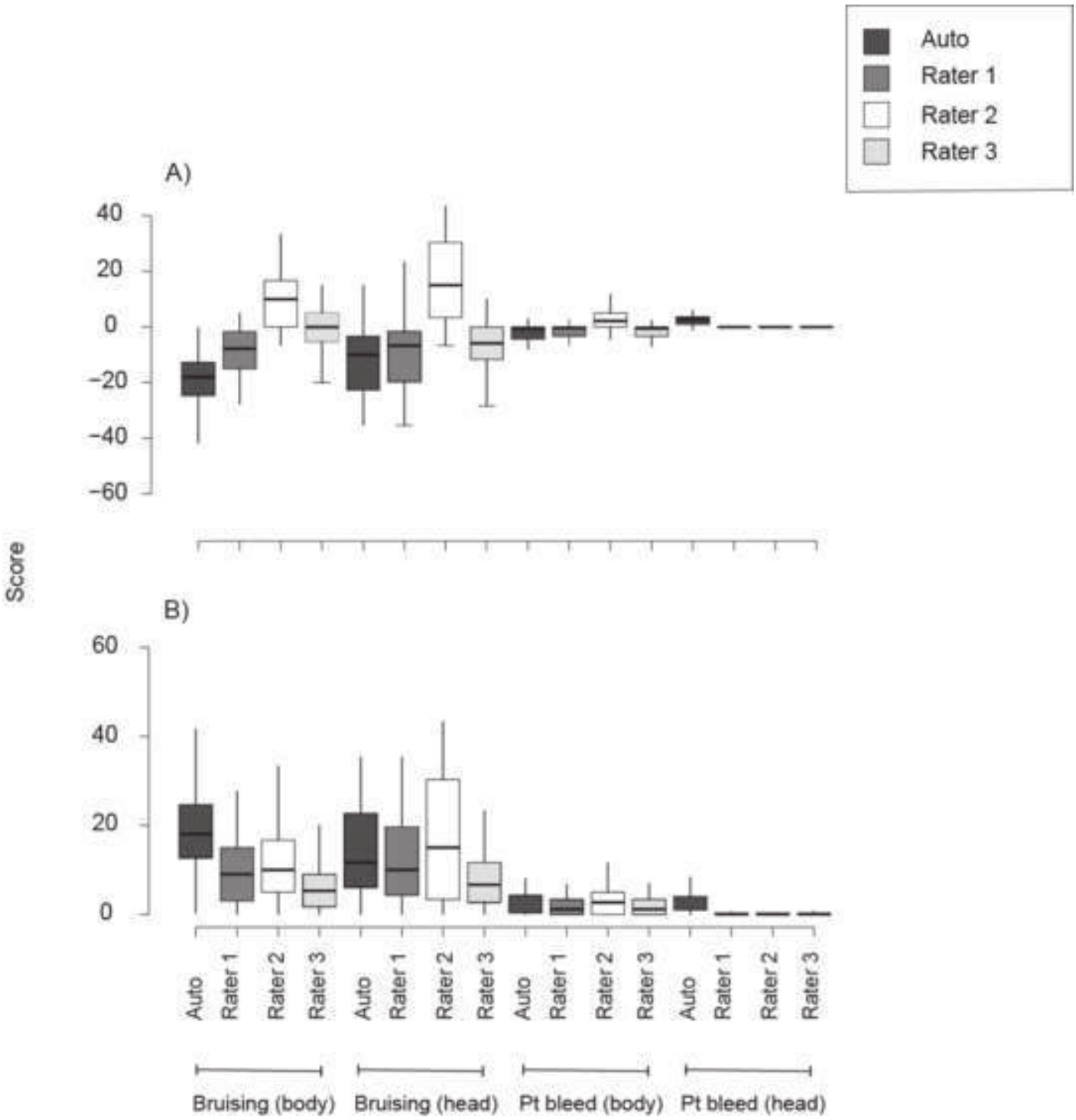

Figure
Click here to download high resolution image